A Kernel-Based Framework for Medical Big-Data Analytics

David Windridge and Miroslaw Bober

Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford, GU2 7XH, UK
{d.windridge,m.bober}@surrey.ac.uk

Abstract. The recent trend towards standardization of Electronic Health Records (EHRs) represents a significant opportunity and challenge for medical big-data analytics. The challenge typically arises from the nature of the data which may be heterogeneous, sparse, very high-dimensional, incomplete and inaccurate. Of these, standard pattern recognition methods can typically address issues of high-dimensionality, sparsity and inaccuracy. The remaining issues of incompleteness and heterogeneity however are problematic; data can be as diverse as handwritten notes, blood-pressure readings and MR scans, and typically very little of this data will be co-present for each patient at any given time interval.

We therefore advocate a kernel-based framework as being most appropriate for handling these issues, using the neutral point substitution method to accommodate missing inter-modal data. For pre-processing of image-based MR data we advocate a Deep Learning solution for contextual areal segmentation, with edit-distance based kernel measurement then used to characterize relevant morphology.

Keywords: Knowledge Discovery, Kernel-Based Methods, Medical Analytics.

1 Introduction

The introduction of electronic health records as a means of standardizing the recording and storage of healthcare information has been significantly accelerating in recent years, resulting in massive volumes of patient data being stored online in a manner readily accessible to clinicians and health-care professionals [1, 2]. Clinicians routinely support patient diagnosis and the selection of the individual treatment by analysis of symptoms in conjunction with the longitudinal patterns evident in physiological data, past clinical events, family history, genetic tests, etc. The availability of this online data resource, covering large cross-sections of the population, offers an unrivaled opportunity to employ big data analytics to spot trends, relations and patterns that may escape the eye of even most experienced clinicians. The results will support personalized medicine, where decisions, practices, and treatments are tailored to the individual patient [3]. (Decision support in particular is a key topic in biomedical

A. Holzinger, I. Jurisica (Eds.): Knowledge Discovery and Data Mining, LNCS 8401, pp. 197–208, 2014.

informatics; it will likely prove essential in clinical practice for human intelligence to be supported by machine intelligence: HumanComputer Interaction and Knowledge Discovery will thus be critical areas of endeavor [4]).

However, analysis of medical records poses several serious challenges due to the nature of the data which is typically heterogeneous, sparse, high-dimensional and often both uncertain and incomplete [5]. Furthermore, image and video data may contain not only the organs/regions of interest but also neighboring areas with little relevance to the diagnosis.

In this paper we propose a kernel-based framework for medical big data analytics to address the issue of heterogeneous data, which employs a *neutral point substitution* method to address the missing data problem presented by patients with sparse or absent data modalities. In addition, since medical records contain many images (X-ray, MRI, etc) we propose to employ a deep-learning approach to address the problem of progressive areal segmentation, in order to improve analysis and classification of key organs or regions.

In the following section we present the kernel-based framework for medical big data analytics. Section 3 presents our arguments for the use of deep learning in medical image segmentation. Final remarks and conclusions are presented in Section 4.

1.1 Glossary and Key Terms

Kernel-Methods: Kernel-Methods constitute a complete machine learning paradigm wherein data is mapped into a (typically) high-dimensional linear feature-space in which classification and regression operations can take place (the support vector machine being the most commonplace). The great advantage of kernel-methods is that, by employing the kernel-trick, the coordinates of the implicitly constructed embedding space need never be directly computed; only the *kernel matrix* of intra-object comparisons are required.

Kernel: A kernel is defined as a symmetric function of arity-2 which forms a positive semidefinite matrix for each finite collection of objects in which pairwise comparison is possible. Critically, this matrix defines a *linear* space such that, in particular, a maximum-margin classifier can be constructed in this embedding space using the (kernelized) support vector machine. The convex optimization problem is solved via the Lagrangian dual, with the decision hyperplane defined by the set of support objects (those with non-zero Lagrangian multipliers).

Neutral Point Substitution: Kernel matrices can be linearly combined while retaining their kernel properties. This makes them ideal for the combination of very heterogeneous data modalities (in multiple-kernel learning the coefficients of this combination are explicitly optimized over). However, the absence of data in any given modality presents significant difficulties in constructing the embedding space. The neutral point substitution method attempts to overcome these in an SVM context by utilizing an appropriate mathematical substitute

to allow embedding-space construction while minimally biasing the classification outcome.

Deep-learning: Deep-learning is the strategy of building multi-layered artificial neural networks (ANNs) via a greedy layer-on-layer process in order to avoid problems with standard back-propagation training. Hinton and Salakhutdinov demonstrated that the greedy layering of unsupervised restricted Boltzmann machines (see below) with a final supervised back-propagation step overcomes many of the problems associated with the deep-layering of tradition ANNs, enabling the building of networks with a distributed, progressively-abstracted representational structure.

Boltzmann machines: A Boltzmann machine consists in a network of units equipped with weighted connection-strengths to other units along with a bias offset. Activation of units is governed stochastically (in contrast to the otherwise similar Hopfield Network) according to the Boltzmann distribution; each unit thus contributes an 'energy' when activated to the overall global energy of the network which derives from the weighted sum of its activations from other units plus the bias. The activation likelihood is itself dependant on the thermodynamic temperature multiplied by this energy magnitude (in accordance with the Boltzmann distribution). Units are themselves split into hidden and visible categories, with the training process consisting in a gradient descent over weights such that the Kullback-Leibler divergence between the thermal equilibrium solution of the network marginalized over the hidden units (obtained via simulated annealing with respect to the temperature) and the distribution over the training set is minimized.

Restricted Boltzmann machines are a special class of the above in which a layering of units is apparent, i.e. such that there are no intra-layer connections between hidden units. These can thus be 'stacked' by carrying out layer-wise training in the manner above, with hidden units from the layer beneath providing the training-data distributions for the layer above.

2 A Kernel-Based Framework for Medical Big Data

Kernel methods [6–8] incorporate important distinctions from traditional statistical pattern recognition approaches, which typically involve an analysis of object clustering within a measurement space. Rather, kernel-based approaches implicitly construct an *embedding space* via similarity measurements between objects, within which the classification (e.g. via an SVM) or regression takes place. The dimensionality of the space is dictated by number of objects and the choice of kernel rather than the underlying feature dimensionality.

Kernel methods thus provide an ideal basis for combining heterogeneous medical information for the purposes of regression and classification, where data can range from hand-written medical notes to MR scans to genomic micro array data. Under standard pattern recognition, pre-processing of individual medical data

modalities would be required to render this data combination problem tractable (i.e. representable in a real vector space), and such representation would invariably involve a loss of information when the data is in a non-vector (e.g graph-based) format.

2.1 Dealing with Heterogeneous Data

Kernel methods provide an excellent framework for combination of heterogeneous data for two principle reasons:

1. Mercer Kernels Are Now for All Data Types
A kernel obeying Mercer's properties (i.e one which leads to a positive definite kernel matrix) can now be built for almost all data formats, for instance:

- **text** (via NLP parse-tree kernels/LSA kernels [9, 10])
- **graphs and sequences** (via string kernels, random walk kernels [11, 12])
- **shapes** (via edit distance kernels [13])
- **real vectors** (via dot product kernels, polynomial kernels etc [14])
- **sets of pixels/voxels** (efficient match kernels, pyramid match kernels [15])
- **stochastic data** (Fisher kernels [8])

Almost all forms of medical data (hand-written medical notes, MR scans, micro array data, temporal blood pressure measurements etc) fall into one or other of these categories. This means that even the most heterogeneous medical data archive is potentially kernalizable.

2. The Linear Combination of a Set of Mercer Kernels Is Also a Mercer Kernel
This means that the combination of kernalizable heterogeneous medical data is straightforward; it is only required to solve the kernel weighting problem (i.e. the optimal coefficients of the linear combination). Fortunately, for the most typical classification context (SVM classification), this is straightforwardly soluble (if no longer convex). Since the sum of kernel matrices of equal size does not increase the summed matrix size we are free (within the limits of the above MKL problem) to add additional kernels indefinitely. This can be useful when multiple kernels are available for each modality, either of different types, or else due to a range a parametric settings available for individual kernels. We can thus capture an enormous range of problem representations - and, of course, we may also employ 'meta' kernels - for instance Gaussian/RBFs built in conjunction with the above kernels, further massively extending the range of possible representations (both individually and collectively). This can be advantageous e.g. for inducing linear separability in the data (Gaussian/RBF kernels are guaranteed to achieve this).

2.2 Dealing with Missing Data

While kernel methods thus, in general, make the problem of combining heterogeneous medical data much more tractable than would be the case for standard pattern recognition methods, there are certain caveats that are specific to

them. The most critical of these is due to the *missing intermodal data problem*, e.g. where a person has incomplete data in a given modality (this is especially common in time-series data where e.g. blood pressure measurements or ECG measurements [16] may have been made irregularly over some interval of the patient's life). This is problematic in standard pattern recognition of course, but can be straightforwardly addressed by interpolating the existing data distribution over the feature space. However, this option is not immediately available for kernel methods, where the data itself *defines* the embedding space.

However, methods developed by the author (the 'neutral point substitution' method [17–19]) render this tractable. (Neutral point substitution involves the unbiased missing value substitution of a placeholder mathematical object in multi-modal SVM combination). We thus, in principle, have the expertise and tool sets required to address the big medical data challenge.

Because of the imaging-based aspect of certain of the data (MR scans in particular), we would anticipate that the kernelized framework for medical data combination would be employed in the context of other computer vision areas, in particular segmentation. Segmentation would be employed, in particular, to identify individual organs prior to kernelized shape characterization for incorporation into the above medical data combination framework.

This aspect of medical segmentation lend itself, in particular, to deep learning approaches, which we now explore.

3 Deep Learning for Medical Image Segmentation

We propose to leverage the hierarchical decompositional characteristics of deep belief networks to address the problem of medical imaging, specifically the aspects of progressive areal segmentation, in order to improve classification of key organs.

Historically, neural networks have been limited to single hidden-layers due to the constraints of back-propagation - specifically the limitations of back-propagation when faced with the parametric freedom and local minima characteristics of the optimization function generated by multiple hidden layers.

The recent development of Deep Networks [20–23] has addressed these issues through the use of a predominantly forward training based approach. Deep networks thus aim to form a generative model in which individual layers encode statistical relationships between units in the layer immediately beneath it in order to maximize the likelihood of the input training data. Thus, there is a greedy training of each layer, from the lowest level to the highest level using the previous layer's activations as inputs. A more recent advancement is the *convolutional deep belief network* [24] that explicitly aims to produce an efficient hierarchical generative model that supports both top-down and bottom-up probabilistic inference; it is these characteristics make it particularly applicable to image processing.

Typically, within a convolutional deep belief network, weights linking the hidden layers to visible layers are distributed over the entire lower layer e.g. an

image pixel grid at the lower layer; in which case the second layer constitutes a bank of image *filters* (note that since higher-level representations are over-complete a sparsity constraint is enforced). However, while such weight learning is a forward-only, greedy, layer-wise procedure, the network's *representation* of an image is constrained by both top-down and bottom-up constraints. Thus, since the network is pyramidal in shape, with *pooling* units serving to enforce compression of representation throughout the network, there is necessarily a bias towards *compositionality* in the network's image representation, since maximizing compositionality and factorizability renders the network as a whole more efficient. Consequently, a convolutional deep belief network is particularly well-suited to capturing progressively higher level hierarchical grammars, with higher levels representing progressively greater abstractions of the data to the extent that the higher levels can embody notions of *categoricity*.

We might, in an ideal case, thus anticipate an encoding of a set of images to consist of the following layers; firstly, an input level of pixel grids; secondly, a set of wavelet-like orthonormal bases for maximal representational compactness over the entire image database; thirdly, a set of feature detectors built from these orthonormal bases but tuned to specific common patterns in the image database. Finally, at the highest levels, we might hope for encoding of broad object categories such as *organs*. Thus, the network as a whole can readily function as an organ segmenter. There is hence a strong continuity of mechanism across the whole representation process, unlike the standard image processing pipeline, in which feature representation and classification are typically separate processes, with attendant numerical mismatches that manifest themselves, for example, as curse-of-dimensionality problems.

All of the above characteristics make deep belief networks and their variants particularly well suited to the proposed application domain of medical imaging. (In particular, a clear hierarchy of image grammars is present; at the highest level there are the individual organs and their relative positional relationship. At the lower level are the organ subcomponents (typically where disease is most manifest). Thus, (in the generative approach) the deep belief network forms a hierarchical segmentation of images in an unsupervised manner, in which different levels of interpretation (i.e. respective conceptual or spatial coarse-grainings) of the data are apparent.

Relevant here are the recent developments by Socher et al. [25] in which a complete NLP grammatical parse tree is distributed across the hierarchy of a recursive deep belief network. In particular, a *syntactically untied recursive neural network* is employed to simultaneously characterize both the relevant phrase structure *and* its most appropriate representation (the system thus learns both the parse tree and a compositional vector representation [i.e a semantic vector space] at the same time). The use of *recursive neural tensor networks* extends this representation to allow more complex 'operator'-type word-classes to exist within the parse tree.

There is no distinction, in principle, between a generative top-down visual grammar of topological relations between segmented regions of a medical image

and the generative (recursive) construction of grammatical units within a NLP parse-tree. Hence, we would anticipate, a recursive, grammatical approach such as the above would be immediately applicable in the medical domain.

We thus, in summary, propose the hierarchical segmentation of medical data by using a deep learning approach. This will require experimentation with the methodology of convolutional deep belief networks in order to optimize the approach for hierarchical image decomposition in a manner most useful to medical objectives. Following this segmentation, morphology and other textural characteristics of the segmented reagion can be treated via an appropriate kernel-characterization in order to allow integration of the deep-learning process into the overall kernel combination framework (using e.g. edit-distance based kernels for contour comparison).

4 Conclusions

In this paper we have outlined two novel research directions to address critical issues in big medical data analytics in a complementary manner: (1) a kernel-based framework for combining heterogeneous data with missing or uncertain elements and (2) a new approach to medical image segmentation built around hierarchical abstraction for later kernel-based characterization.

Kernelization is thus the key to addressing data-heterogeneity issues; the way in which missing 'intermodal' data is combined in within the kernel-based framework depends on the authors' *neutral point substitution method*. A neutral point is defined as a unique (not necessarily specified) instantiation of an object that contributes exactly zero information to the classification (where necessary to actuate this substitution explicitly -i.e. where missing data occurs in both test and training data- we can select the minimum norm solution)

This is therefore an ideal substitute for missing values in that it contributes no overall bias to the decision. Crucially, it can be used in multi-kernel learning problems, enabling us to combine modalities optimally with arbitrary missing data.

The calculation of the neutral points turns the $O(n^3)$ complexity of the SVM problem into a maximum $O(n^4)$ complexity problem if uniform intermodal weightings are used. Solving for arbitrary weights, in the general MKL optimization procedure, is inherently a non-convex problem and solved via iterative alternation between maximizing over Lagrangian multipliers and minimizing over the modality weights. Appropriate modality weightings can thus be learned in multimodal problems of arbitrary data completion; we hence have the ability to combine any multiple modality data irrespective of modal omission.

Together, the two approaches of kernel-based combination of heterogeneous data with neutral point substitution and deep-learning would thus, we argue, address the major outstanding big-data challenges associated with Electronic Health Record standardization, in particular incompleteness and heterogeneity. However, there remain certain outstanding issues to be addressed:

5 Open-Problems

While the proposed kernel-based framework is extremely generic, in that almost any form of data can be accommodated, there has typically been a historical bias in kernel research towards classification (particularly SVM-based classification), particularly in terms of the algorithms that have deployed for kernel-based analysis. Classification, however, accounts for only a fraction of the medical big-data activity that one would wish to carry-out (in particular, it corresponds to the activity of *disease diagnosis*, though perhaps classification could also be employed for the induction of key binary variables relating to health-care, such as determining whether a patient is likely to be a smoker given the available evidence). The first open problem that we can thus identify is that of the *1. Breadth of Problem Specification.*

A second issue that we would need to address in an extensive treatment of EHRs is that of *2. Data Mining/Unsupervised Clustering Analysis.* Here, we wish to utilize all the available data (in particular the exogenous variables) in order to suggest investigative possibilities, rather than explicitly model or classify patient data. This approach thus differs from standard modes of assessment in which one seeks to test a null hypothesis against the available data. Typically, data mining is thus a precursory stage to hypothesis evaluation; instead it belongs to the stage of *hypothesis generation.* We may thus envisage a 'virtuous circle' of activity with progressive problem specification arising from the iterative interaction between medical practitioners and machine-learning researchers, with medical hypothesis suggestions being following by the progressive formalization of diagnostic evaluation criteria. The hypothesis suggestions themselves would arise from unsupervised clustering analysis; significant multimodality would be suggestive of subpopulations within the data, in particular subpopulations that may not be recognized within existing disease taxonomies.

A third related open problem is that of *3. Longitudinal Data Analysis* for both patients and any identified subpopulations within the data. This would generally take the form of *prediction modelling*, in which we would attempt to determine to what extent it is possible to predict an individual patient's disease prognosis from the data. A sufficiently comprehensive data set that extends across the diverse range of medical measurement modalities would potentially enable novel forms of analysis; for instance, apparently exogenous variables could prove to affect outcomes in different ways.

One additional aspect of time-series data that would also potentially have to be addressed is that of on-line learning; patient data would, in general, be collected continuously, and therefore methods of classification and regression would preferably therefore have to be trainable incrementally i.e. with relatively little cost involved in retraining with small quantities of additional data.

The final open problem associated with our framework is that of *4. Utilizing Human-Computer Interaction* in the most effective way, particularly as regards decision support. Here, we would not attempt to to directly resolve problems of e.g. medical diagnosis/prognosis via regression/classification. Rather, the aim is to utilize the framework to maximally assist the medical profession in arriving

at their own diagnosis/prognosis decision. This might take the form of data-representation i.e where machine learning is used to determine salient aspects of the data set with respect to the current decision to be made. This can be at both the high-level (determination of patient or population context) or at the low-level (segmentation of relevant structures within imaging data). Another possibility for HCI is explicit hybridization of the decision process [4], utilizing the most effective respective areas of processing in combination (for example, pattern-spotting in humans in conjunction with large data-base accessing in computers).

In the following section we outline some provisional solutions and research directions for addressing the most immediately tangible of these open issues consistent with the proposed framework.

6 Future Outlook

To address the first open problem, *Breadth of Problem Specification*, it will likely be necessary, in addition to carrying-out classification within a kernel context, to exploit the full range of algorithms to which kernel methods apply: *kernel-principal component analysis*, *ridge regression*, *kernel-linear discriminant analysis*, *Gaussian processes* etc (these are typically all convex optimization problems). The latter of these, Gaussian processes, ought particularly to be useful for prediction of patient outcomes [26], and would directly assist with the missing data problem by allowing both longitudinal data-interpolation and longitudinal data-extrapolation.

A Gaussian process might thus be deployed for modelling time-series data via Gaussian process regression (kriging), producing the best linear unbiased prediction of extrapolation points. Primary outcome variables in a medical context would thus be likely to be disease progression indicators such as tumor-size.

We therefore propose to explore a number of areas of kernel regression consistent with our framework to address the big-data analytics problem.

In terms of the second open problem *Data Mining/Unsupervised Clustering Analysis*, we might wish to explore rule-mining type approaches, since the resulting logical clauses most closely resemble the diagnostic criteria used by medical professionals in evaluation disease conditions.

Unsupervised clustering, in particular the problem of determining the presence of sub-populations within the data, might be addressed by addressed by model-fitting (Kernel regression [27], or perhaps K-means or Expectation Maximization with Gaussian Mixture Modelling in sufficiently low-dimensional vector spaces) in conjunction with an appropriate model-selection criterion (for example, the Akaike Information Criterion [28] or Bayesian Information Criterion). The latter is required to correctly penalize model-parameterization with respect to goodness-of-fit measurements such that the overfitting-risk is minimized. The means of such partitioned clusters would then correspond to *prototypes* within the data, which may be used for e.g. efficient indexing and kernel based characterization of novel data. Manifold learning might also be required in order

to determine the number of active factors within a data-set; Gaussian Process Latent Variable Modelling [29] would be a good fit to the proposed framework.

More generally, regarding the use of such hypothesis generation methods within a medical context, it is possible can regard the iterative process of experimental feedback and problem refinement as one of *Active Learning* [30]. Active-learning addresses the problem of how to choose the most informative examples to train a machine learner; it does so by selecting those examples that most effectively differentiate between hypotheses, thereby minimizing the number of experimental surveys that need to be conducted in order to form a complete model. It thus represents a maximally-efficient interaction model for medical professions and machine-learning researchers.

Finally, addressing the third open problem, *Longitudinal Analysis*, within the proposed kernel-based framework would likely involve the aforementioned Gaussian processes given their ready incorporation within the kernel methodology. Another possibility would be Structured Output Learning [31], a variant of the Support Vector Machine that incorporates a loss-function capable of measuring the distance between two structured outputs (for example, two temporal series of labels). The Structured Output Learning problem, however, is generally not tractable with standard SVM solvers due to the additional parameter complexity and thus requires bespoke iterative cutting plane-algorithms to train in polynomial time.

In sum, it would appear that the proposed kernel framework has sufficient flexibility to address many of the open questions identified, and indeed implicitly sets out a programme of research to address these. However, it will invariably be the case that each novel EHR dataset will involve characteristics that are more suited to one particular form of machine learning approach over another -it is not generally the case that this can be specified *a priori*, thereby necessitating a flexible research programme with the potential to leverage the full range of available kernel-based machine-learning techniques.

References

1. Holzinger, A.: Biomedical Informatics: Discovering Knowledge in Big Data. Springer, New York (2014)
2. Holzinger, A., Dehmer, M., Jurisica, I.: Knowledge discovery and interactive data mining in bioinformatics state-of-the-art, future challenges and research directions. BMC Bioinformatics 15(suppl. 6)(I1) (2014)
3. Simonic, K.M., Holzinger, A., Bloice, M., Hermann, J.: Optimizing long-term treatment of rheumatoid arthritis with systematic documentation. In: Proceedings of Pervasive Health - 5th International Conference on Pervasive Computing Technologies for Healthcare, pp. 550–554. IEEE (2011)
4. Holzinger, A.: Human–computer interaction & knowledge discovery (hci-kdd): What is the benefit of bringing those two fields to work together? In: Cuzzocrea, A., Kittl, C., Simos, D.E., Weippl, E., Xu, L. (eds.) CD-ARES 2013. LNCS, vol. 8127, pp. 319–328. Springer, Heidelberg (2013)

5. Marlin, B.M., Kale, D.C., Khemani, R.G., Wetzel, R.C.: Unsupervised pattern discovery in electronic health care data using probabilistic clustering models. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, IHI 2012, pp. 389–398. ACM, New York (2012)
6. Scholkopf, B., Smola, A.: MIT Press (2002)
7. Shawe-Taylor, J., Cristianini, N.: Cambridge University Press (2004)
8. Hofmann, T., Schlkopf, B., Smola, A.J.: A review of kernel methods in machine learning (2006)
9. Collins, M., Duffy, N.: Convolution kernels for natural language. In: Advances in Neural Information Processing Systems 14, pp. 625–632. MIT Press (2001)
10. Aseervatham, S.: A local latent semantic analysis-based kernel for document similarities. In: IJCNN, pp. 214–219. IEEE (2008)
11. Nicotra, L.: Fisher kernel for tree structured data. In: Proceedings of the IEEE International Joint Conference on Neural Networks IJCNN 2004. IEEE press (2004)
12. Borgwardt, K.M., Kriegel, H.P.: Shortest-path kernels on graphs. In: Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM 2005), pp. 74–81. IEEE Computer Society, Washington, DC (2005)
13. Daliri, M.R., Torre, V.: Shape recognition based on kernel-edit distance. Computer Vision and Image Understanding 114(10), 1097–1103 (2010)
14. Smola, A.J., Ovri, Z.L., Williamson, R.C.: Regularization with dot-product kernels. In: Proc. of the Neural Information Processing Systems (NIPS), pp. 308–314. MIT Press (2000)
15. Grauman, K., Darrell, T.: The pyramid match kernel: Efficient learning with sets of features. J. Mach. Learn. Res. 8, 725–760 (2007)
16. Holzinger, A., Stocker, C., Bruschi, M., Auinger, A., Silva, H., Gamboa, H., Fred, A.: On applying approximate entropy to ecg signals for knowledge discovery on the example of big sensor data. In: Huang, R., Ghorbani, A.A., Pasi, G., Yamaguchi, T., Yen, N.Y., Jin, B. (eds.) AMT 2012. LNCS, vol. 7669, pp. 646–657. Springer, Heidelberg (2012)
17. Panov, M., Tatarchuk, A., Mottl, V., Windridge, D.: A modified neutral point method for kernel-based fusion of pattern-recognition modalities with incomplete data sets. In: Sansone, C., Kittler, J., Roli, F. (eds.) MCS 2011. LNCS, vol. 6713, pp. 126–136. Springer, Heidelberg (2011)
18. Poh, N., Windridge, D., Mottl, V., Tatarchuk, A., Eliseyev, A.: Addressing missing values in kernel-based multimodal biometric fusion using neutral point substitution. IEEE Transactions on Information Forensics and Security 5(3), 461–469 (2010)
19. Windridge, D., Mottl, V., Tatarchuk, A., Eliseyev, A.: The neutral point method for kernel-based combination of disjoint training data in multi-modal pattern recognition. In: Haindl, M., Kittler, J., Roli, F. (eds.) MCS 2007. LNCS, vol. 4472, pp. 13–21. Springer, Heidelberg (2007)
20. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Computation 1, 541–551 (1989)
21. Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy layer-wise training of deep networks. In: Schölkopf, B., Platt, J., Hoffman, T. (eds.) Advances in Neural Information Processing Systems 19, pp. 153–160. MIT Press, Cambridge (2007)
22. Ranzato, M., Huang, F.J., Boureau, Y.L., LeCun, Y.: Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: CVPR. IEEE Computer Society (2007)

23. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science 313(5786), 504–507 (2006)
24. Lee, H., Grosse, R., Ranganath, R., Ng, A.Y.: Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations (2009)
25. Socher, R., Manning, C.D., Ng, A.Y.: Learning continuous phrase representations and syntactic parsing with recursive neural networks
26. Shen, Y., Jin, R., Dou, D., Chowdhury, N., Sun, J., Piniewski, B., Kil, D.: Socialized gaussian process model for human behavior prediction in a health social network. In: 2012 IEEE 12th International Conference on Data Mining (ICDM), pp. 1110–1115 (December 2012)
27. Song, C., Lin, X., Shen, X., Luo, H.: Kernel regression based encrypted images compression for e-healthcare systems. In: 2013 International Conference on Wireless Communications Signal Processing (WCSP), pp. 1–6 (October 2013)
28. Elnakib, A., Gimel'farb, G., Inanc, T., El-Baz, A.: Modified akaike information criterion for estimating the number of components in a probability mixture model. In: 2012 19th IEEE International Conference on Image Processing (ICIP), pp. 2497–2500 (September 2012)
29. Gao, X., Wang, X., Tao, D., Li, X.: Supervised gaussian process latent variable model for dimensionality reduction. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 41(2), 425–434 (2011)
30. Cai, W., Zhang, Y., Zhou, J.: Maximizing expected model change for active learning in regression. In: 2013 IEEE 13th International Conference on Data Mining (ICDM), pp. 51–60 (December 2013)
31. Yan, J.F., Kittler, Mikolajczyk, K., Windridge, D.: Automatic annotation of court games with structured output learning. In: 2012 21st International Conference on Pattern Recognition (ICPR), pp. 3577–3580 (November 2012)